A Novel Feature Selection Algorithm using Multi-Objective Improved Honey Badger Algorithm and Strength Pareto Evolutionary Algorithm-II

DOI : 10.36909/jer.16863

Anusha Papasani*, Nagaraju Devarakonda
School of Computer Science and Engineering VIT-AP University, Vijayawada, India
* Corresponding Author: anoosha.papasani@gmail.com

ABSTRACT

An important task for classification is feature selection that removes the redundant or irrelevant features from the dataset. Multi-objective feature selection approach is mainly proposed by many researchers. However, these approaches failed to maintain the higher classification accuracy while removing redundancy in the features. In this work, a wrapper based feature selection technique is proposed with a hybrid of Multi Objective Honey Badger Algorithm (MO-HBA) and Strength Pareto Evolutionary Algorithm-II to maintain the balance between classification accuracy and removal of redundancy. Classification accuracy improvement and removal of redundant features are considered as the multi-objective optimization functions of the proposed multi-objective feature selection technique. The Levy flight algorithm is utilized to initialize the population to enhance the ability of the exploration and exploitation of MO-HBA. The regularized Extreme Learning Machine is used to classify the selected features. To evaluate the performance of the proposed feature selection technique, eighteen benchmark datasets are utilized and results are compared with the four well known multi-objective feature selection techniques in terms of accuracy, hamming loss, ranking loss, mean value, standard deviation, length of features, and training time. The proposed approach achieved maximum accuracy of 100% with the maximum value of selected features as 80. The minimum value of hamming loss, ranking loss, mean value and standard deviation value achieved by the proposed approach are 0.0092, 0.0003, 0.018 and 0.001 respectively. The experimental results show that the proposed approach can give improved classification accuracy while the removal of redundancy in large scale datasets.

Keywords: Strength Pareto Evolutionary Algorithm-II; Multi-objective optimization; Wrapper feature selection; Levy flight; Honey badger algorithm.

INTRODUCTION

In data mining and machine learning, the main issue is classification. With most relevant features, collection and creation of dataset is a hard task. But the collected features are redundant and irrelevant to the classifiers (Hu et al., 2020). To get a more compact dataset, selection of maximum relevant features is necessary. One of effective dimensionality reduction techniques to solve this issue is feature selection (Xue et al., 2021). With the accuracy improvement, the feature selection technique enables classifiers to obtain better interpretability, improves the classifier's generalization ability and reduces over-fitting (Dash et al., 1997). Filters and wrappers are the two main categories of the feature selection algorithms. Using the intrinsic data characteristics, the key features are selected by the Filter methods. Based on roughest theory, distance and information theory the important features selected by the Filter methods (Labani et al., 2020). In the wrapper method, learning algorithms are used to evaluate the significance of selected features. Based on the various search strategies, the important features are selected by the Wrapper methods (Xue et al., 2013). In the classification performance, the wrapper method is better than the filter methods (Vijayanand et al., 2020). But the time consumption of wrapper methods for large scale datasets is more than the filter methods (Bermejo et al., 2014). For minimizing the size of feature subset and improving the accuracy of classification, from an optimization point of view the wrapper feature selection can be termed as a multi-objective optimization model (Li, et al., 2020). Due to the powerful searching ability in the total workspace, a lot of attention is gained by meta-heuristic optimization techniques (Nayyar et al., 2018, Zhang et al., 2018 & Le et al., 2018). So that, they have been broadly used in several real-world applications like path planning of unmanned aerial vehicle, flow shop scheduling problem and wireless sensor networks. In wrapper algorithm, the meta-heuristic optimization algorithm plays an important role. Particle Swarm Optimization (PSO), Bat algorithm (BA), Whale optimization algorithm (WOA), Genetic algorithm (GA) and Grey Wolf Optimization (GWO) algorithm are the widely used optimization algorithms. In which, PSO (Banka et al., 2015 & Xue et al., 2014) and GA (Eroglu et al., 2017 & Hamdani et al., 2011) are widely used in many studies. Based on the number of objective functions, these evolutionary algorithms are further divided into multi-objective methods and single-objective methods (Han et al., 2015). When compared to single-objective methods, the multi-objective methods have more advantages (Fu et al., 2014). Though the generally used single-objective optimization algorithms such as PSO and GA achieve better global search performance, their exploitation ability of identified regions is weak. Honeybadger Algorithm (HBA) (Hashim et al., 2022) is a recently proposed optimization algorithm among these algorithms with better exploration and exploitation ability. The multi-objective methods such as Strength Pareto based evolutionary algorithm-II (SPEA-II) (Zitzler et al., 2001), non-dominated sorting genetic algorithm II (NSGA-II) (Deb et al., 2002) and Pareto Envelope-based Selection Algorithm (PESA) (Knowles et al., 1999). have shown better performance in handling many objective functions. Among them, SPEA-II is the popular and very important evolutionary algorithm that is the improved version of SPEA. It is the best multi-objective framework when compared with NSGA-II and PESA, particularly in high-dimensional spaces. The feature selection aimed at minimizing error rate while the removal of redundant features. Thus the multi-objective optimization techniques are considered more in the feature selection approaches. The existing multi-objective optimization algorithms such as SPEA-II, NSGA-II and PESA produce non-dominated solutions and show lesser accuracy with the reduction in the feature size (Dong et al., 2020). The combination of many optimization algorithms can improve the classification accuracy while the reduction in the feature size. In this work, the combination of HBA with SPEA-II named MOHBSP2 is proposed as the multi-objective optimization algorithm to select the features. The Levy flight algorithm (Liu et al., 2020) is utilized to generate the initial population of MOHBSP2. This method can extend the search space and can improve the performance of existing SPEA-II approach with guarantee in search speed. Due to the imbalanced dataset, the performance of classifier will be get affected. To balance the datasets, the mostly used effective sampling method is Synthetic minority over-sampling technique (SMOTE) (Huang et al., 2020). In the proposed work, the preprocessing step involves the SMOTE technique to balance the dataset. To calculate the accuracy of the selected features, a classifier is used by the proposed wrapper technique. Many efficient and accurate classification algorithms such as Naive Bayes, Decision tree, SVM, Artificial Neural network and Convolutional Neural networks are utilized by many researchers to classify the selected features. Among them Extreme Learning Machine (ELM) is the good and fast learning algorithm. In this work Regularized Extreme Learning Machine (RELM) (Wang et al., 2020) is employed to evaluate the performance of proposed approach. When compared to conventional classifiers, RELM can give satisfactory results in classification of large-scale datasets with multi-label. The results of proposed method is compared with popular feature selection algorithms such as multi-objective binary genetic algorithm integrating an adaptive operator selection mechanism (MOBGA-AOS) (Xue et al., 2021), Multi-objective PSO (MO-PSO) (Paul et al., 2021), multi-objective binary cuckoo optimization algorithm and non-dominated sorting genetic algorithms NSGA III (BCNSG3) (Usman et al., 2020) and modified whale optimization algorithm (MWOA) (Vijayanand et al., 2020).

The main contributions of the proposed MOHBSP2 based feature selection technique can be summarized as follows: The HBA is used for the multi-objective feature selection algorithm. The SPEA-II algorithm is combined with the MO-HBA named MOHBSP2 to improve the performance of existing HBA. To improve the exploration and exploitation ability of existing SPEA-II approach, the Levy flight is used to initialize the population. The effectiveness of the proposed feature selection approach is evaluated by utilizing the selected features in classification. The selected features are used to improve the classification performance of existing Regularized Extreme learning Machine. The performance of the proposed MOHBSP2 based feature selection algorithm is measured in terms of accuracy, number of selected features, computation time, hamming loss, ranking loss mean value, and standard deviation, to show the effectiveness of proposed feature selection algorithm. The performance of proposed approach is compared with other four well-known multi-objective evolutionary algorithms such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO on eighteen datasets.

RELATED WORKS

Most of Feature selection methods are proposed as a single-objective problem in the fitness function. Multi Objective feature selection methods are only few. For multi-objective feature selection, Zhang et al. (2020) introduced a binary differential evolution based self-learning. They have used three operators such as purifying search one-bit operator, probability difference binary mutation and non-dominated sorting operator based on crowding distance. This approach is evaluated using twenty standard datasets. This approach showed better performance and consumes shorter running time. With two objective functions, wrapper based multi-objective feature selection technique based on NSGA-II is proposed by Kozodoi et al. (2019) In the objective functions, feature subset reduction and expected profit maximizing are considered. They have utilized ten retail credit scoring datasets. This suggested approach showed better performance. Using breeding operators and NSGA-II, a wrapper based multi-objective feature selection approach is proposed by Gonzalez et al. (2019). This method maintains the stability with the feature ranking process. They have applied four classifiers to evaluate the effectiveness of the proposed approach. With the combination of salp swarm algorithm and multi-objective spotted hyena optimizer a feature selection approach is introduced by Sharma and Rani (2019). There are two stages in this method. The irrelevant features are eliminated using filter approach in the first stage. Then the most optimal features are explored using their hybrid method in the second stage. Kiziloz et al. (2018) proposed teaching learning based multi-objective feature selection approach. In this approach three multi-objective TLBO algorithms such as non-dominated selection, scalar transformation and minimum distance are proposed. They have utilized three classifiers namely SVM, ELM and LR for the performance evaluation of the proposed approach. Based on the frequency in the set of documents, the features are ranked using a PSO based multi-objective approach proposed by Amoozegar and Minaei-Bidgoli (2018). Then, the particles are guided and set of archives are enhanced using these levels. This proposed approach is compared with multi-objective GA and three variants of PSO. The hybrid of non-dominated sorting genetic operators and multi-objective artificial bee colony is introduced by Hancer et al. (2018) for feature selection. This approach used the k nearest neighbor classifier to evaluate the selected feature subset. For the classification of microarray results and gene selection, a multi-objective approach is proposed by Dashtban et al. (2018) They have used SVM, NBY, KNN and DT as the classifiers to evaluate the effectiveness of proposed approach. For gene selection, hybrid filter-wrapper based multi-objective optimization approach is introduced by Lai (2018). They select the finest genes using an aggregate filter technique. In a multi-objective formula, contribution lies and elitism based differential evolution approach is proposed by Nayak et al. (2020) for filter based feature selection. This approach removes the undesired and redundant features of the processed dataset by considering the nonlinear and linear dependencies. Usman et al. (2020) proposed a filter based feature selection algorithm using multi objective binary cuckoo optimization algorithm and NSGA-III. They have proposed four multi-objective filter based feature selection techniques by utilizing entropy based on gain ratio and mutual information. Though their method derives best feature subsets, performance comparison with other evolutionary algorithms is not provided to show their effectiveness. Xue et al. (2021) introduced a multi-objective feature selection technique for classification using binary genetic algorithm and adaptive crossover operator. In their approach, different search characteristics are used with five crossover operators. According to the performance of the evolution process, a probability is assigned to each of the crossover operators. A Relative Discriminative Criterion (MORDC) algorithm with multi objective function is suggested by Labani et al. (2020) for text feature selection. They have taken the relevance of the text features to the target class as the first objective and evaluation of connection between the features is selected as the second objective. They compute the redundancy and relevancy of the features using Pearson correlation and RDC measures. Gao et al. (2021) introduced a multi-objective optimization algorithm to select the features with hybrid cat swarm optimization algorithm (HCSO). In this approach, they have combined the inherent, competitive and guided characteristics with the original CSO. They have evolved global worst and best solutions during the HCSO execution. Though this approach gives better results, the comparison results with other existing optimization algorithms are not provided. An improved multi-objective PSO based feature selection approach for medical datasets is proposed by Rostami et al. (2020). They have used a mutation operator to enhance the quality of generated feature subset and the diversity of BPSO. Moreover, convergence of the PSO algorithm improved by introduction of a new node centrality-based approach to optimize the initial population of PSO. Hybrid of Harris Hawks Optimization and Fruitfly Optimization Algorithm is proposed by Abdollahzadeh and Gharehchopogh (2021). In this multi-objective approach, the error rate is considered as one objective function and number of features is considered as another objective function. The performance of these existing approaches showed lesser accuracy with the reduction in the feature size. Though the existing multi-objective optimization algorithms namely SPEA-II showed better performance than other approaches, it produces non-dominated solutions and shows lesser accuracy with the reduction in the feature size. From this literature review, it can be known that, the combination of optimization approaches can give better results in feature selection. So that, hybrid of various optimization algorithms is proposed in this paper to improve the performance of existing approaches such as SPEA-II and HBA.

PRELIMINARIES

Honey Badger Algorithm

HBA is inspired by the honey badger's intelligent foraging behavior (Hashim et al., 2022). Algorithm 1 explains the pseudo code for Honey Badger Algorithm. Into the exploitation and exploration phases in HBA, the honey badger's search behavior is formulated with the honey finding and digging techniques. Representation of the population (P) of the candidate solution is given in Equation (1).

$$P = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1D} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2D} \\ & & \dots\dots\dots\dots\dots\dots. & & \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nD} \end{bmatrix} \tag{1}$$

The ith position of honey badger is represented by using Equation (2).

$$a_i = [a_i^1, a_i^2, \dots, a_i^D] \tag{2}$$

Algorithm 1. Pseudocode for Honey Badger Algorithm

Initialize parameters t_{max}, N, β, C
Generate population P_{new} with random position between 0 and 1.
Evaluate fitness of each position x_i using objective function f_i
The best position x_{prey} and the fitness is assigned to f_{prey}
while t≤ t_{max} **do**
    Update the decreasing factor ∝
    for i=1 to N **do**
        Calculate the intensity I_i
        if r<0.5 **then**
          Update the position x_{new}
        else
          Update the position x_{new}
        end if
        Evaluate new position and assign to f_{new}
        if $f_{new} \leq f_i$ **then**
            Set x_i=x_{new} and $f_i = f_{new}$
        end if
        if $f_{new} \leq f_{prey}$ **then**
            Set x_{prey}=x_{new} and $f_{prey} = f_{new}$
        end if
  end for
end while Stop criteria satisfied
Return x_{prey}

Strength Pareto Evolutionary Algorithm-II (SPEA-II)

SPEA-II is one of the evolutionary multiple objective algorithms (Zitzler et al., 2001). For the multiple objective optimization problems, the original genetic algorithm is extended as SPEA-II. To preserve and identify a set of Pareto optimal solutions is the objective of this algorithm. The Pareto optimal set is all the Pareto optimal solutions. In the objective space, the best non-dominated solutions made the Pareto optimal set. For each solution, the two main parameters are considered. The algorithm 2 details the pseudocode for SPEA-II.

Algorithm 2. Pseudocode for SPEA-II

Population (P) initialization
Creation of empty external set E
    for i=1 **to** no of Generation
        Calculation of fitness for each features in A and P
        Add Non-dominated features from A and P
        if capacity of A exceeds **then**
            By Truncation Operator Remove features from A
        if capacity of A not exceeds **then**
            To fill E using dominated features in P
        Binary tournament selection
        Mutation and crossover
  end for

PROPOSED METHOD

The proposed approach selects the features using MOHBSP2 approach and the selected features are classified using regulated extreme learning machine. The existing multi-objective approaches select the non-dominated parameters and the accuracy is decreasing with the reduction in the feature size. Improving the accuracy of classification process while the reduction in the feature size is the complex task. The combination of more optimization algorithms can solve this issue. In this paper, the performance of the existing SPEA-II approach is improved with the combination of HBA and Levy flight (Balakrishnan et al., 2021, Ewees et al., 2022) approaches. Thus the proposed approach can give better accuracy while the reduction in the feature size.

Dataset

Eighteen datasets from UCI Machine Learning Repository (Asuncion and Newman, 2007) are utilized for evaluating the performance of the proposed feature selection approach. The datasets utilized to evaluate the proposed approach are WDBC, Zoo, Lymphography, Ionosphere, Credit, Heart, Dermatology, Sonar, Spect, Parkinson, Indian Pima, Scene, Kc1, Audiology, Tic-tac-toe, Waveform, Glass and Wine collected from various domains such as medical, computer.

MOHBSP2 Based Feature Selection

According to the basic HBA, a modified multi-objective method called MOHBSP2 is proposed in this paper. The exploitation ability can be improved by the local search with Levy Flight.

Initialization

In this phase, the parameters like maximum iteration count, archive size ($\overline{N}$) and feature size are initialized. At t = 0, the initial population(P_0) is generated. The size of population and respective positions are initialized based on Equation (3)

$$x_i = lb_i + r_1 \times (ub_i - lb_i) \tag{3}$$

Where r_1 represents a random number between 1 and 0. The ith position of honey badger referring to feature in a population P_0 is represented by x_i The lower and upper bounds of the search domain are represented by lb_i and ub_i. Then an empty archive $\overline{(P_t)}$ is initialized by Levy flight algorithm using Equation (4).

$$\overline{P}_t = x_i + \propto \oplus Levy(\lambda) \tag{4}$$

Where parameter of random step size is represented by $\propto$, the distribution parameter of Levy flight is represented by λ. The entry wise multiplication is represented by $\oplus$. The ith solution is represented by x_i.

Fitness Assignment

The fitness values are calculated for both the population P_t and $\overline{P}_t$. A strength value S(i) is allocated by each individual i in the population P_t and archive $\overline{P}_t$. It denotes the dominated number of features as expressed in Equation (5).

$$\mathrm{S(i)} = |\{\mathrm{j}| j \in P_t + \overline{P}_t \wedge i\ \phi\ j)| \tag{5}$$

Where, the cardinality of a set is represented by $|\cdot|$. The Pareto dominance relation is represented by ϕ. The raw fitness Rw(i) of a feature is calculated on the basis of the S values as given in Equation (6).

$$Rw(i) = \sum_{j \in P_t + \overline{P_t}, j \phi i} S(j) \tag{6}$$

Corresponding to i, the density D(i) is expressed as given in Equation (7),

$$\text{D (i)} = \frac{1}{\sigma_i^k + 2} \tag{7}$$

Where, the k-th element for each individual i is represented by σ_i^k. For the raw fitness value Rw(i) the addition of D(i) produces its fitness Ft(i) as given in Equation (8).

$$\text{Ft(i)} = \text{Rw(i)} + \text{D(i)} \tag{8}$$

Fitness Function

For finding a set of optimal features with small solution size and high classification accuracy is the aim of the multi-objective feature selection technique. Instead of maximizing the classification accuracy, minimization of classification error is taken as the first fitness function. The second fitness function considers the size of solution. To evaluate the solutions, k nearest neighbors (k-NN) is utilized as the classifier. The k-NN uses n-fold cross validation. The first fitness function can be calculated using Equation (9).

$$\min(f_1) = \left(\frac{1}{n} \sum_{l=1}^{n} \frac{N_{Error}}{N_{All}}\right) \times 100\% \tag{9}$$

Where the feature is represented by X, The number of all the instances are represented by N_{All}, the number of wrongly predicted instances are represented by N_{Error}. The second fitness function can be calculated using Equation (10),

$$\min(f_2) = \sum_{i=1}^{D} x_i \tag{10}$$

Where, ith value in the X feature is represented by x_i. The number of original features is represented by D.

Environmental Selection

All the non-dominated features from both the populations $\overline{P_t}$ and P_t that have fitness value lower than one are copied to $\overline{P_{t+1}}$. It can be expressed by using Equation (11),

$$\overline{P_{t+1}} = \{i | i \in P_t + \overline{P_t} \wedge F(i) < 1 \tag{11}$$

The environmental selection is completed if the front of non-dominated fits exactly into the dataset $|\overline{P_{t+1}}| = \overline{N}$. If the size of dataset is too small ($\overline{P_{t+1}} | < \overline{N}$) then, $\overline{P_{t+1}}$ fill with dominated features from $\overline{P_t}$ and P_t. The truncation operator is used to reduce $\overline{P_{t+1}}$ if the size of dataset is too large ($\overline{P_{t+1}} | > \overline{N}$).

Position Update using HBA

Multi-objective HBA is utilized to update the positions of population $\overline{P_{t+1}}$. The positions of $\overline{P_{t+1}}$ is updated using the Equations (12) if the value of r is less than 0.5 otherwise Equation (13) can be used.

$$\overline{P_{t+1}} = x_{prey} + F \times \beta \times I \times x_{prey} + F \times r_3 \times \alpha \times d_i \times |\cos(2\pi r_4) \times [1 - \cos(2\pi r_5]| \tag{12}$$

$$\overline{P_{t+1}} = x_{prey} + F \times r_7 \times \propto \times\, d_i \tag{13}$$

Where, the global best position is represented by x_{prey}. The honey badger's ability to search food is represented by β. The distance between prey and i^{th} honey badger is represented by

d_i and it can be calculated using Equation (14). The random numbers between 0 and 1 are represented by r_3, r_4, r_5, r_7. The search direction is represented by F and it can be calculated using Equation (15). The decreasing factor α is updated using Equation (16).

$$d_i = x_{prey} - x_i \tag{14}$$

Where, the global best position is represented by x_{prey} and i^{th} honey badger position is represented by x_i.

$$\text{F} = \begin{cases} 1 & if\ r_6 \le 0.5 \\ -1 & else \end{cases} \tag{15}$$

The random number between 0 and 1 is represented by r_6.

$$\alpha = C \times \exp\left(\frac{-t}{t_{max}}\right) \tag{16}$$

Where, t_{max} represents the maximum count of iterations. The constant C have default value as 2.

Feature selection using SPEA-II

If the maximum iteration is reached then the non-dominated features are represented by A with the set of decision vectors. To fill the mating pool, with the replacement of $\overline{P_{t+1}}$ binary tournament selection is performed. For the mating pool, the mutation and crossover operators are applied and resulting population is $\overline{P_{t+1}}$. Then the generation counter is incremented to t = t+1. Then the fitness is calculated.

Computational Complexity

The computational complexity of the proposed approach is O(m n^2) where the population size is represented by n and the number of selected features is represented by m.

System Model

The proposed feature selection approach is established with following steps. Step 1: The collected datasets are preprocessed using scaling function. Step 2: The high dimensionality of the preprocessed dataset is reduced using the proposed MOHBSP2 based feature selection approach. Step 3: Then the selected features are classified using the RELM. The proposed approach showed better accuracy with reduced feature size and computational time. At first, the parameters used in the proposed approach are initialized. The population is initialized with random numbers between 0 and 1. Then archive set is generated by using levy flight algorithm. The fitness value is evaluated for each population. Based on the fitness value the positions are updated using Equations (12-13). Thus the position update is performed with HBA. Then selection, crossover and mutation process are performed based on SPEA-II. Based on these values the optimal features are selected from the dataset. The selected features are used in the RELM based classification.

Algorithm 3. Pseudo code for proposed feature selection algorithm

Initialize parameters t_{max}, N, β, C, $\bar{N}$
Generate new population P_0 with random position between 0 and 1
Generate archive set $\overline{P_0}$ (t=0) based on Levy flight using Equation (4) **// Levy flight**
while t≤ t_{max} **do**
    Update the decreasing factor ∝ using Equation (16)
    for i=1 to Number of populations **do**
        Calculate fitness of each feature in P_t and $\overline{P_t}$
        $\overline{P_{t+1}}$ = Copy all non-dominated feature from $\overline{P_t}$ and $\overline{P_t}$ to $\overline{P_{t+1}}$
        if r<0.5 **then**
            Update the position $\overline{P_{t+1}}$ using Equation (12) **// HBA**
        else
            Update the position $\overline{P_{t+1}}$ using Equation (13)
        end if
    end for
Perform selection **//SPEA-II**
Apply crossover and mutation
end while stop criteria satisfied
Calculate and save feature subset's error rate (solutions)
Using non-dominated sorting ranking the population
Return the selected features
end

EXPERIMENTAL RESULTS

In the Figure 1, the non-dominated solutions are plotted for the evaluation of performance of the proposed approach on feature subset searching. The proposed approach result is taken for Zoo dataset and compared with existing approaches. The size of the feature subset is represented by the horizontal axis of the graph and classification error is represented by the vertical axis of graph. From the pareto front results it can be known that, the proposed MOHBSP2 approach can achieve lower classification error while ensuring the smaller solution size for all the eighteen datasets.

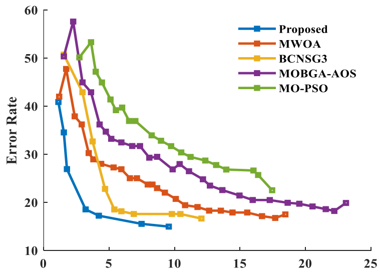
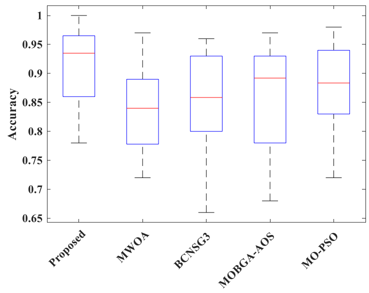
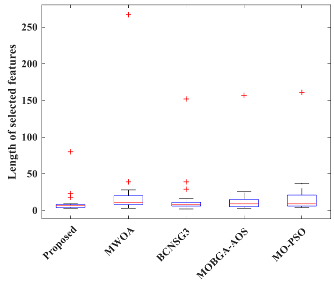


Figure 1. Pareto front analysis

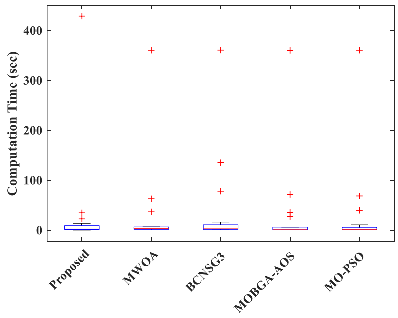
In Figure 2(a) the accuracy of the proposed approach is compared with existing approaches in terms of box plot analysis for all the datasets. For the proposed approach, the datasets ionosphere, Parkinson and wine have achieved 100% accuracy. The upper quartile in the box plot for the proposed approach is 100%. For the existing approaches such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO algorithms the highest value of upper quartile is 97%, 96%, 97% and 98% respectively. So that, the proposed approach has highest value of upper quartile than existing approaches in the box plot. The median line of the box for the proposed approach is at 93% that is higher than the existing approaches. There are no potential outliers in the box plot, as all the approaches has better average performance. Similarly, the selected feature length is compared with existing approaches in terms of box plot analysis in Figure 2(b). The highest length of selected features in the proposed approach is 80. For the existing approaches such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO algorithms the highest value of selected features is 267, 152, 157 and 161 respectively. Due to the large size of dataset, there are some potential outliers in the box plot for all the approaches. Thus proposed approach selects less number of features than the existing approaches. The computation time taken by the proposed approach is compared with existing approaches in terms of box plot analysis in the Figure 2(c). The highest value of computation time taken by the proposed approach is 378.99 seconds. For the existing approaches such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO algorithms the highest value of computation time is 350.58 seconds, 328.82 seconds, 337.23 seconds and 365.59 seconds respectively. The computation time taken by the proposed approach is 13.31 seconds higher than the existing approaches. The difference in the computation time is in seconds. From these results it can be known that, the proposed feature selection approach can maintain higher accuracy with the reduction of feature size.



(a) Accuracy



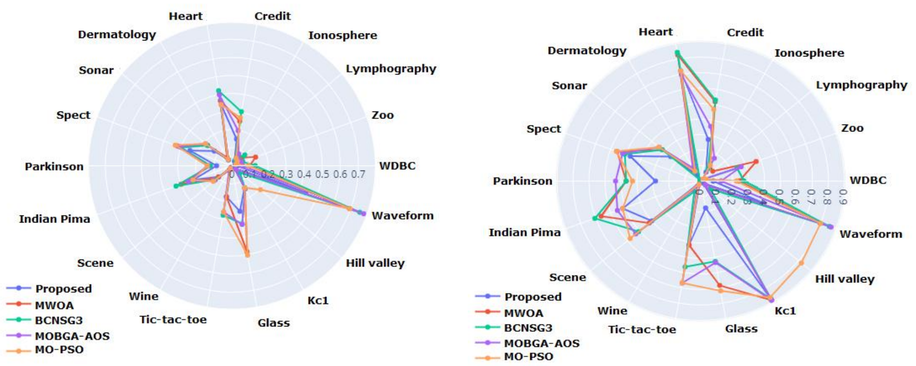
(b) Feature length



(c) Computation time

Figure 2. Box plot analysis of accuracy

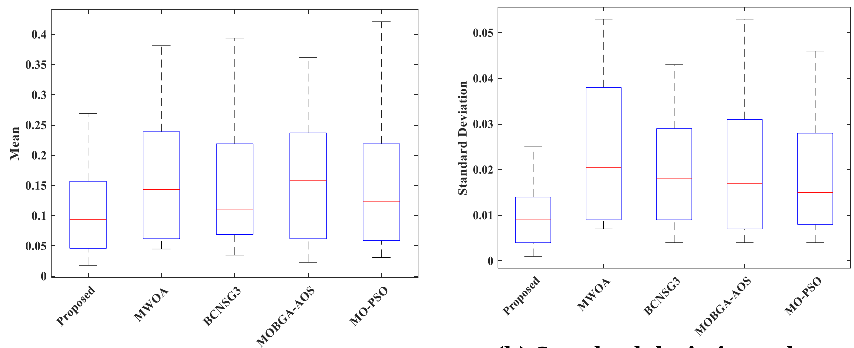
In the Figure 3(a) and (b), the hamming loss and ranking loss results of the proposed feature selection approach are compared with other four well known feature selection approaches such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO using radar chart. The minimum value of hamming loss of the proposed approach is 0.0092. For the existing approaches such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO the lowest value of hamming loss in the box plot is 0.0294, 0.0198, 0.0268, and 0.0153 respectively. The minimum value of ranking loss for the proposed approach is 0.0003. For the existing approaches such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO the lowest ranking loss is 0.0029, 0.0020, 0.0027, and 0.0015 respectively. For the better performance of feature selection approach, the values of hamming loss and ranking loss should be less. From the radar charts it can be known that the proposed feature selection approach has less hamming loss and ranking loss when compared with other techniques.



(a) Hamming Loss **(b) Ranking Loss**

Figure 3. Radar chart for Hamming Loss and Ranking Loss

The performance of the proposed feature selection algorithm is evaluated using IGD metric in terms of standard deviation and mean value for all eighteen datasets. The Figure 4 shows the box plot analysis of mean value and standard deviation. The results of proposed MOHBSP2 method is compared with other multi-objective feature selection algorithms such as MOBGA-AOS, MO-PSO, BCNSG3, and MWOA. From these results it can be known that the proposed feature selection approach has less standard deviation and mean values for all the eighteen datasets when compared with other algorithms. The performance comparison for the proposed MOHBSP2 approach is compared with other existing multi-objective algorithms such as MWOA, BCNSG3, MOBGA-AOS and MO-PSO for all the eighteen datasets. The proposed approach has highest value of accuracy when compared to other four algorithms for all the eighteen datasets. Moreover, the proposed approach selects less number of features than other algorithms. Thus, the proposed approach minimizes the classification error while minimizing the number of selected features. But the training time taken by the proposed approach is higher than other approaches. From these results it can be concluded that the proposed multi-objective feature selection technique can give better results than other conventional multi-objective feature selection techniques.



(a) Mean value **(b) Standard deviation value**

Figure 4. Box plot analysis for mean value and standard deviation value

CONCLUSION

A novel multi-objective wrapper feature selection technique is proposed in this paper to select less number of features with less classification error. Hybrid of Multi Objective Honey Badger Algorithm and Strength Pareto Evolutionary Algorithm-II with Levy flight algorithm named MOHBSP2 is the proposed feature selection algorithm. The effectiveness of the proposed approach is evaluated using eighteen benchmark datasets and compared with four well known multi-objective feature selection techniques. The Regularized extreme learning machine is utilized to evaluate the selected features using proposed feature selection approach. From the classification results, it can be known that, the proposed feature selection technique outperforms other feature selection approaches with less number of selected features and less classification error. The proposed approach achieved 100% accuracy with the maximum length of selected features is 80. The minimum value of hamming loss, ranking loss, mean value and standard deviation value achieved by the proposed approach are 0.0092, 0.0003, 0.018 and 0.001 respectively. Though the proposed approach shows better performance in terms of accuracy, ranking loss and hamming loss, the time cost of proposed approach is slightly larger than other approaches. The reduction of computation complexity of the proposed multi-objective feature selection algorithm will be investigated in the future work.

REFERENCES

Abdollahzadeh, B., & Gharehchopogh, F. S. 2021. A multi-objective optimization algorithm for feature selection problems. Engineering with Computers 1-19.

Amoozegar, M., & Minaei-Bidgoli, B. 2018. Optimizing multi-objective PSO based feature selection method using a feature elitism mechanism. Expert Systems with Applications 113: 499-514.

Asuncion, A., & Newman, D. 2007. UCI machine learning repository.

Balakrishnan, K., Dhanalakshmi, R. & Khaire, U.M. 2021. Improved salp swarm algorithm based on the levy flight for feature selection. The Journal of Supercomputing 77(11): 12399-12419.

Banka, H., & Dara, S. 2015. A Hamming distance based binary particle swarm optimization (HDBPSO) algorithm for high dimensional feature selection, classification and validation. Pattern Recognition Letters 52: 94-100.

Bermejo, P., Gámez, J. A., & Puerta, J. M. 2014. Speeding up incremental wrapper feature subset selection with Naive Bayes classifier. Knowledge-Based Systems 55: 140-147.

Dash, M., & Liu, H. 1997. Feature selection for classification. Intelligent Data Analysis 1:131–156.

Dashtban, M., Balafar, M., & Suravajhala, P. 2018. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. Genomics 110(1): 10-17.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. 2002. A fast and elitist multi objective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation 6(2): 182-197.

Dong, H., Sun, J., Sun, X., & Ding, R. 2020. A many-objective feature selection for multi-label classification. Knowledge-Based Systems 208.

Eroglu, D. Y., & Kilic, K. 2017. A novel Hybrid Genetic Local Search Algorithm for feature selection and weighting with an application in strategic decision making in innovation management. Information Sciences, 405: 18-32.

Ewees, A. A., Mostafa, R. R., Ghoniem, R. M., & Gaheen, M. A. 2022. Improved seagull optimization algorithm using Lévy flight and mutation operator for feature selection. Neural Computing and Applications 1-36.

Fu, W., Xue, B., & Zhang, M. 2014. Multi-objective feature selection in classification: A differential evolution approach. In Asia-Pacific Conference on Simulated Evolution and Learning. Springer, Cham.

Gao, X. Z., Nalluri, M. S. R., Kannan, K., & Sinharoy, D. 2021. Multi-objective optimization of feature selection using hybrid cat swarm optimization. Science China Technological Sciences 64(3): 508-520.

Gonzalez, J., Ortega, J., Damas, M., Martín-Smith, P., & Gan, J. Q. 2019. A new multi-objective wrapper method for feature selection–Accuracy and stability analysis for BCI. Neurocomputing 333: 407-418.

Hamdani, T. M., Won, J. M., Alimi, A. M., & Karray, F.2011. Hierarchical genetic algorithm with new evaluation function and bi-coded representation for the selection of features considering their confidence rate. Applied Soft Computing 11(2): 2501-2509.

Han, M., & Ren, W. 2015. Global mutual information-based feature selection approach using single-objective and multi-objective optimization. Neurocomputing 168: 47-54.

Hancer, E., Xue, B., Zhang, M., Karaboga, D., & Akay, B. 2018. Pareto front feature selection based on artificial bee colony optimization. Information Sciences 422: 462-479.

Hashim, F. A., Houssein, E. H., Hussain, K., Mabrouk, M. S., & Al-Atabany, W. 2022. Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems. Mathematics and Computers in Simulation 192:84-110.

Hu, P., Pan, J. S., & Chu, S. C. 2020. Improved Binary Grey Wolf Optimizer and Its application for feature selection. Knowledge-Based Systems 195.

Huang, L., Zhang, H., Wu, C. Q., & Li, Z. 2020. An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset. Computer Networks 177.

Kiziloz, H. E., Deniz, A., Dokeroglu, T., & Cosar, A. 2018. Novel multiobjective TLBO algorithms for the feature subset selection problem. Neurocomputing 306: 94-107.

Knowles, J., & Corne, D. 1999. The Pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimisation. Proceedings of the 1999 Congress on Evolutionary Computation, 98–105.

Kozodoi, N., Lessmann, S., Papakonstantinou, K., Gatsoulis, Y., & Baesens, B. 2019. A multi-objective approach for profit-driven feature selection in credit scoring. Decision support systems 120: 106-117.

Labani, M., Moradi, P., & Jalili, M. 2020. A multi-objective genetic algorithm for text feature selection using the relative discriminative criterion. Expert Systems with Applications 149.

Lai, C. M. 2018. Multi-objective simplified swarm optimization with weighting scheme for gene selection. Applied Soft Computing 65: 58-68.

Le, D. N., Nayyar, A., & Nguyen, N. G. 2018. Advances in swarm intelligence for optimizing problems in computer science. CRC press.

Li, A. D., Xue, B., & Zhang, M. 2020. Multi-objective feature selection using hybridization of a genetic algorithm and direct multisearch for key quality characteristic selection. Information Sciences 523: 245-265.

Liu, Y., & Cao, B. 2020. A novel ant colony optimization algorithm with Levy flight. IEEE Access 8: 67205-67213.

Nayak, S. K., Rout, P. K., Jagadev, A. K., & Swarnkar, T. 2020. Elitism based multi-objective differential evolution for feature selection: A filter approach with an efficient redundancy measure. Journal of King Saud University-Computer and Information Sciences 32(2): 174-187.

Nayyar, A., & Nguyen, N. G. 2018. Introduction to swarm intelligence. Advances in swarm intelligence for optimizing problems in computer science 53-78.

Paul, D., Jain, A., Saha, S., & Mathew, J. 2021. Multi-objective PSO based online feature selection for multi-label classification. Knowledge-Based Systems 222.

Rostami, M., Forouzandeh, S., Berahmand, K., & Soltani, M. 2020. Integration of multi-objective PSO based feature selection and node centrality for medical datasets. Genomics 112(6): 4370-4384.

Sharma, A., & Rani, R. 2019. C-HMOSHSSA: Gene selection for cancer classification using multi-objective meta-heuristic and machine learning methods. Computer methods and programs in biomedicine 178: 219-235.

Usman, A. M., Yusof, U. K., & Naim, S. 2020. Filter-based multi-objective feature selection using NSGA III and cuckoo optimization algorithm. IEEE Access 8: 76333-76356.

Vijayanand, R. & Devaraj, D. 2020. A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network. IEEE Access 8: 56847-56854.

Wang, D., Wang, P., Yuan, Y., Wang, P., & Shi, J. 2020. A fast conformal predictive system with regularized extreme learning machine. Neural Networks 126: 347-361.

Xue, B., Zhang, M., & Browne, W. N. 2013. Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach. IEEE Transactions on Cybernetics 43: 1656–1671.

Xue, B., Zhang, M., & Browne, W. N. 2014. Particle swarm optimization for feature selection in classification: Novel initialisation and updating mechanisms. Applied soft computing 18: 261-276.

Xue, Y., Zhu, H., Liang, J., & Słowik, A. 2021. Adaptive crossover operator based multi-objective binary genetic algorithm for feature selection in classification. Knowledge-Based Systems 227.

Zhang, X., Zhang, Q., Chen, M., Sun, Y., Qin, X., & Li, H. 2018. A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method. Neurocomputing 275: 2426-2439.

Zhang, Y., Gong, D. W., Gao, X. Z., Tian, T., & Sun, X. Y. 2020. Binary differential evolution with self-learning for multi-objective feature selection. Information Sciences 507: 67-85.

Zitzler, E., Laumanns, M., & Thiele, L. 2001. SPEA2: Improving the strength pareto evolutionary algorithm, EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems. Athens. Greece.